

# Digital Libraries: Re-inventing Scholarly Information Dissemination and Use

*Robert Wilensky*

*Principal Investigator*

*David Forsyth*

*Co-principal Investigator*

*The UC Berkeley Digital Library Team*



# Who we are

---

- ◆ PI and Co-PI:
  - Robert Wilensky (CS & SIMS)
  - David Forsyth (CS)
- ◆ Faculty Investigators
  - Richard Fateman (CS)
  - Marti Hearst (SIMS)
  - Joe Hellerstein (CS)
  - James Landay (CS)
  - Ray Larson (SIMS)
  - Jitendra Malik (CS)
  - Philip Stark (Statistics)
  - Robert Twiss (CED)
  - Doug Tygar (CS & SIMS)
  - Nancy Van House (SIMS)
  - Hal Varian (SIMS)
- ◆ Post-docs
  - Chad Carson
  - Tom Phelps
- ◆ Other Investigators
  - Henry Baird (Xerox PARC)
  - Bernie Hurley (UCB Library)
- ◆ Students
  - Serge Belongie
  - Hao Chen
  - Byunghoon Kang
  - Thomas Leung
  - Taku Tokuyasu
  - Barbara Rosario
  - Shengdong Zhao
- ◆ Staff
  - Ginger Ogle
  - Jeff Anderson-Lee
  - Howard Foster
  - Loretta Willis
  - Joyce Gross
  - Tony Morosco



# Who we plan to work with

---

## ◆ UCB Organizations

- U.C.B. Library
- U.C.B. Instructional Technology Program
- Museum of Vertebrate Zoology
- UCB School of Environmental Science Policy and Management

## ◆ Corporate

- Xerox PARC
- Hewlett-Packard
- NEC
- SUN Microsystems
- IBM Almaden
- Microsoft
- Sharp

## ◆ *DLIB InterOp Project* Partners

- Stanford, UCSB
- California Digital Library
- SDSC

## ◆ Not-for-profits

- California Academy of Science
- California Department of Fish and Game
- California Native Plant Society
- UC Davis ICE project
- Museum Digital Library Collection
- USGS
- Federal Bureau of Investigation
- others tba



# Current (Print) System

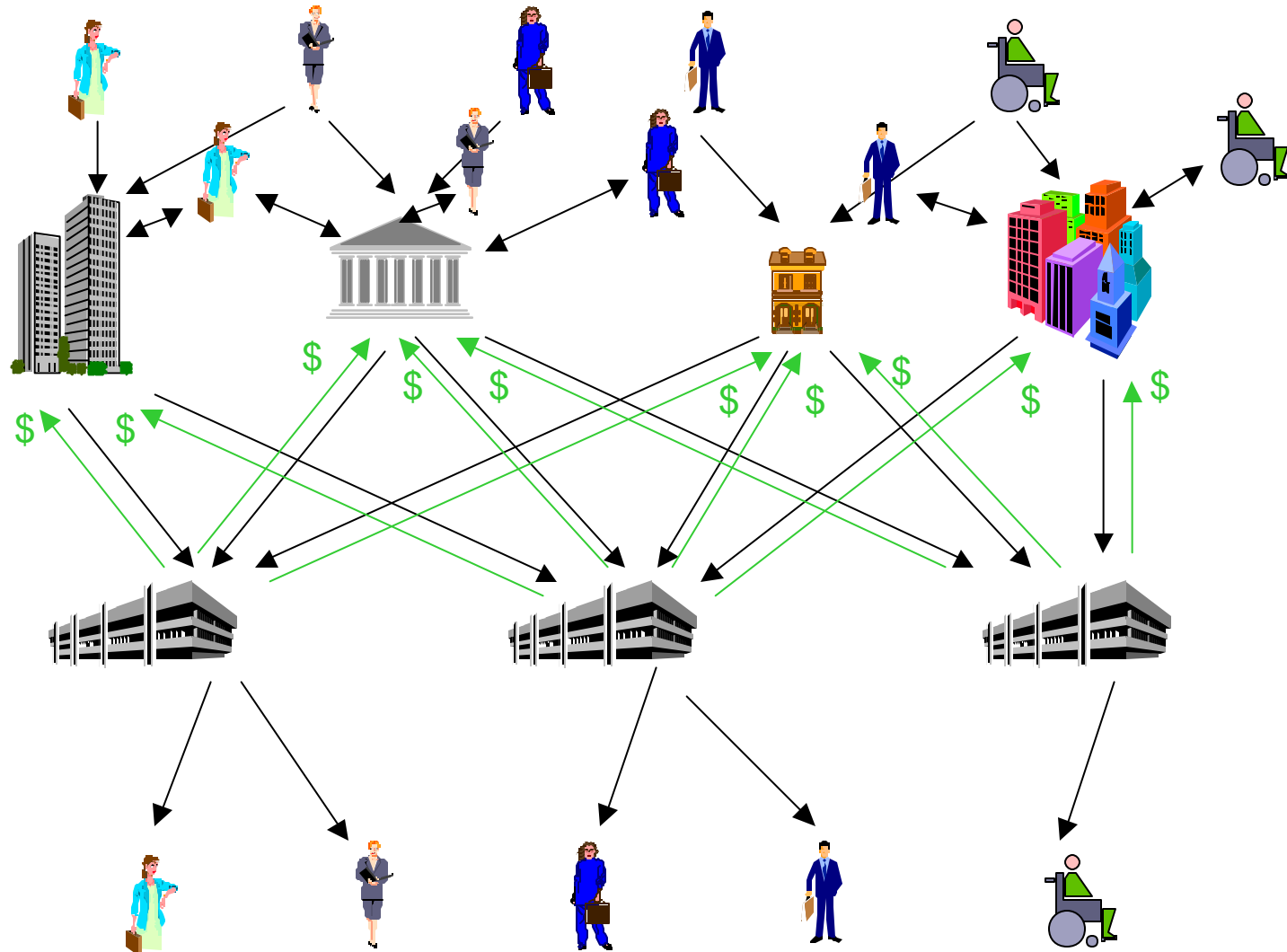
Originators

Reviewers

Publishers

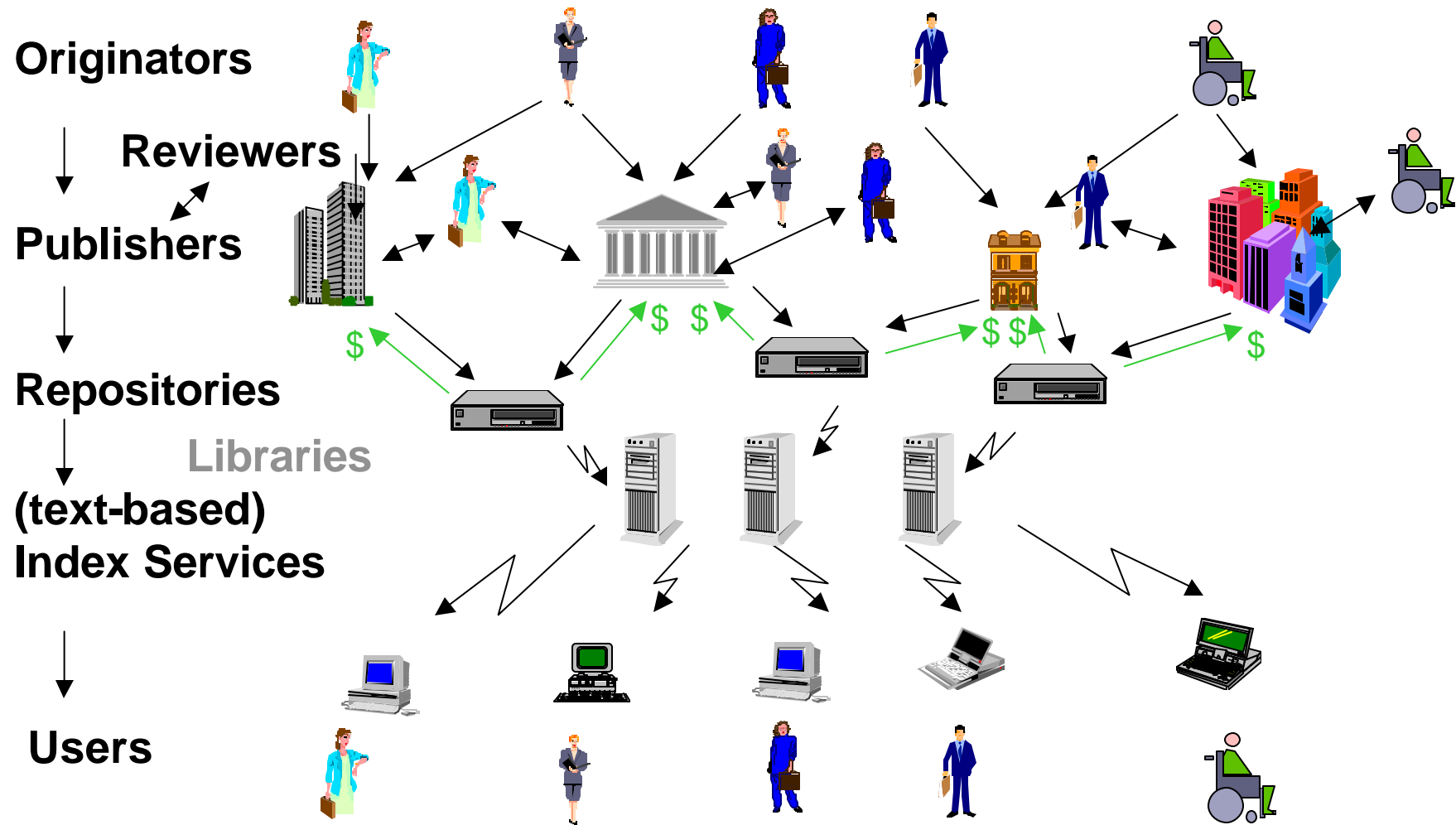
Libraries

Users



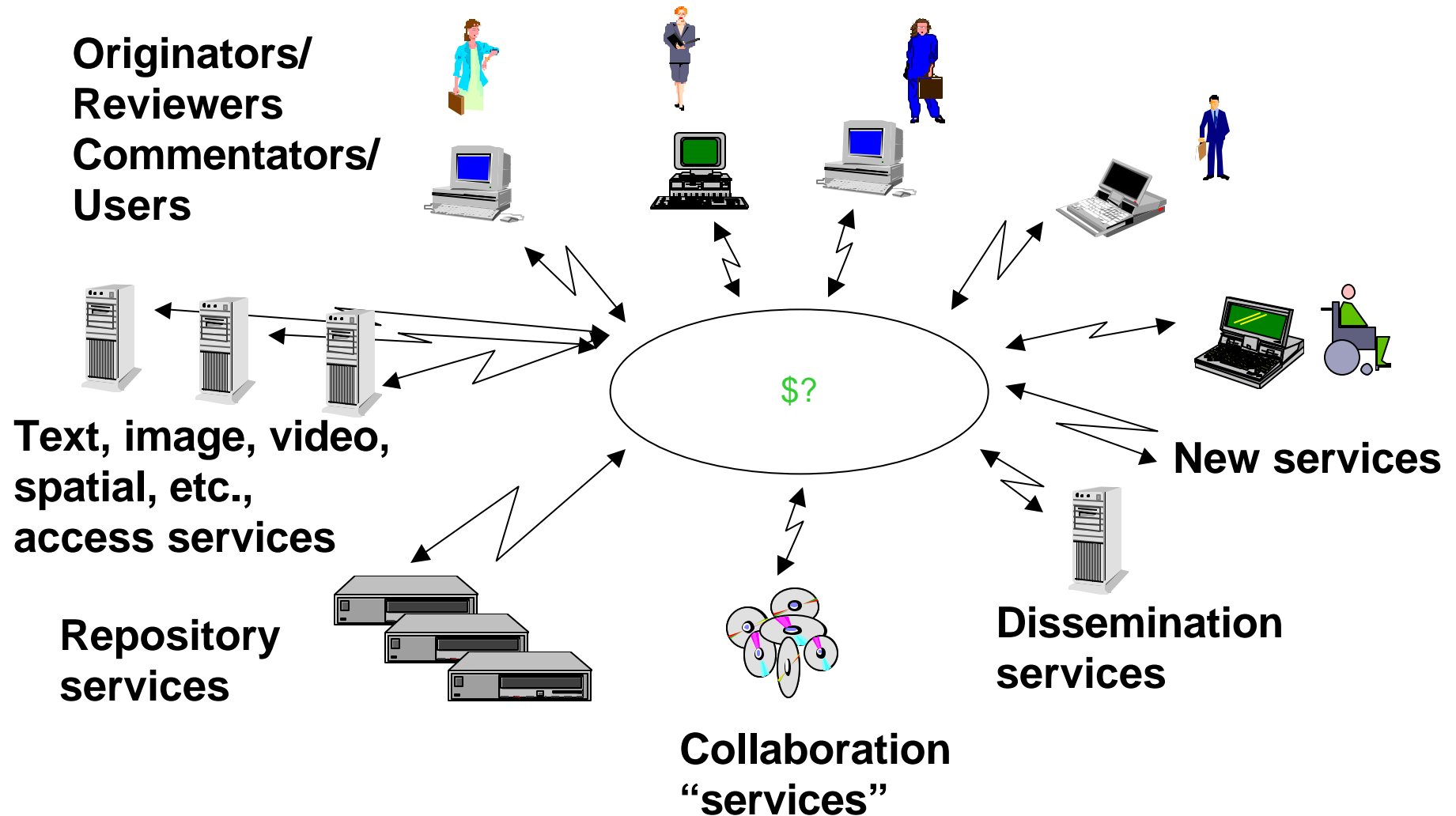


# Current Electronic System





# A Fully Distributed Model





# What we'd like to make it possible for users to do

---

- ◆ Easily find, access and annotate any available digital object (journal paper, tech report, data set, map, annotation) in any format.
  - to be used by peer reviewers, NSF reviewers, colleagues, students
- ◆ Spontaneously make a group of distributed users and give them access privileges to a paper/annotation/datum.
- ◆ Find all the annotations that have been made on one's draft/published paper.
- ◆ Click on a citation in a scanned document image, generating a search for the underlying reference.
- ◆ Access federated repositories that contain all relevant data for a scientific purpose; contrast against other information
  - E.g., all occurrences of a given species in North America, presented as a county-by-county distribution map.
- ◆ Incorporate papers, notes, annotations, etc. into an everywhere available electronic notebook.
- ◆ Benefit from what your colleagues have found useful.
- ◆ Have system that will support all of the above and is both economically and technologically scalable.



# Desirable technologies

---

- ◆ Collection and dissemination
  - easy to pass information to a (virtual) collection
- ◆ Search
  - potential users must be able to find what they want
- ◆ Viewing
  - clean display of complex datasets
- ◆ Interaction and annotation
  - by referees, collaborators, others





# Collection example

---

- ◆ InterBiome (I.e., CalFlora++)
  - Basis for a federated set of national on-line, in-depth, freely available information resource of American life forms
  - Technological enablers
    - ◆ general taxon name translation service.
    - ◆ location resolution service
  - Experiment in alternative information dissemination:
    - » information continually disseminated by originators
    - » critiqued by distributed experts
    - » federated with similar and complementary resources
  - Collaboration with many partners



# Testbed Data Status

Type	Examples	Sep 99	
Documents	articles, EIRs, water reports	280,237 pp.	67 GB
Images	DWR Cal. Flora Corel Animals, etc. Total	17,601 20,286 39,100 1,875 78,862	474GB
Aerial photos	Suisun March Sac-SJ Delta	1074 img	3.4GB
Sensor Data	Delta fish flow	30 days	.02MB
GIS Data	dams, fish, watersheds, etc.	various	50MB
DOQs	SF Bay Area	219 img	33GB
DRGs	California		26GB
CalFlora DB	Occurrences	674,814	539MB
Other tables	MVD, streets	1,285,096	185GB
<b>Total</b>			<b>789GB</b>



## Collections - the future

---

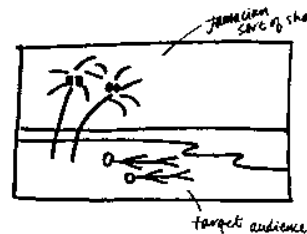
- ◆ Work with MVZ to expand biological collections
- ◆ Work with ECAI to obtain tools for art collection
- ◆ Respond to other opportunities - we didn't see Calflora coming!



# Content Analysis for Access

---

- ◆ Image/video Analysis
- ◆ Document Analysis
- ◆ Text Analysis



We've got the picture.



We've got the picture.



Telephone: 0171 367 1146 Fax: 0171 752 9305



Telephone: 0171 367 1146 Fax: 0171 752 9305



# Image Analysis

---

- ◆ **Blobworld implementation**
  - probably best “general” image finding-by-image-content technology
- ◆ **Body plans**
  - much higher accuracy for specific object classes
  - requires lots of engineering per object class



# Blobworld Status

Two side-by-side screenshots of a Netscape browser window displaying "Blobworld Query Results: image #23018 (Prefiltered)".

The left screenshot shows a query image of a person in a white outfit and a corresponding blob. Below it is a table titled "blob and feature importance:".

	blob (overall)	color	texture	location	shape
blob 5	very	very	somewhat	somewhat	somewhat

Below the table, it says "Querying from 35000 images (2000 returned by the filter)".

The right screenshot shows a grid of 16 query results, each with an image, a score, and a "New query" link.

- 7. 348038 (score = 0.96) [New query](#)
- 8. 298075 (score = 0.96) [New query](#)
- 9. 105028 (score = 0.96) [New query](#)
- 10. 286099 (score = 0.96) [New query](#)
- 11. 271045 (score = 0.95) [New query](#)
- 12. 105040 (score = 0.95) [New query](#)
- 13. 123067 (score = 0.95) [New query](#)
- 14. 358030 (score = 0.95) [New query](#)
- 15. 304005 (score = 0.95) [New query](#)
- 16. 280074 (score = 0.95) [New query](#)





# Blobs vs terms

Query image: 84079

Query blob

Querying from 35000 images (2000 returned by the filter).

blob and feature importance:					
	blob (overall)	color	texture	location	shape
blob 1	very	very	somewhat	not	somewhat

1: 84068 (score = 0.96) [New query](#)

2: 84072 (score = 0.91) [New query](#)

3: 35148 (score = 0.9) [New query](#)

4: 221011 (score = 0.9) [New query](#)

5: 132011 (score = 0.87) [New query](#)

6: 84022 (score = 0.87) [New query](#)

Netscape: Query results using keywords only

File Edit View Go Communicator Help

Querying from 35000 images 216 have the keyword "rose."

1: 8112 [New query](#)

2: 8122 [New query](#)

3: 11085 [New query](#)

4: 13061 [New query](#)

5: 24012 [New query](#)

6: 20049 [New query](#)

7: 24026 [New query](#)



8: 35155 [New query](#)



# Blobs AND terms

Netscape: Blobworld Query Results: image #84079 (Keyword Match)



File Edit View Go Communicator Help





blob and feature importance:					
	blob (overall)	color	texture	location	shape
blob 1	very	very	somewhat	not	somewhat

Query image: 84079      Query blob



Querying from 35000 images (216 have the keyword "rose.")





1: [84068](#) (score = 0.96)      [New query](#)



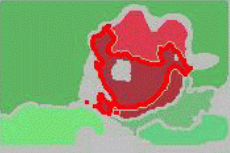

2: [84072](#) (score = 0.91)      [New query](#)





3: [84022](#) (score = 0.87)      [New query](#)



4: [84016](#) (score = 0.84)      [New query](#)



5: [84037](#) (score = 0.81)      [New query](#)



6: [84093](#) (score = 0.81)      [New query](#)





# Content Analysis - the future

---

- ◆ Handle such queries as
  - “which flowers are easily confused” (CALFLORA)
  - “how often do religious pictures show...” (ECAI)
- ◆ by developing
  - a new framework for segmentation
    - » e.g., checked petals
  - reasoning about spatial arrangements of regions
    - » e.g., fritillary=many checked petals in a ring
  - reasoning about specific activities
    - » e.g., person with arms above head = striking a blow



# Distributed Geographic Information Use

---

- ◆ Objective: Investigate geodata federating technologies.
- ◆ Approach: Develop proxy services that use OGDl to access OpenGIS-compliant (and OGDl) repository servers.
- ◆ UCB GIS Viewer extended to provide data visualization services
- ◆ InterLib collaboration:
  - Data discovery and access enhanced as proposed by UCSB
  - OGDl geodata transformation via a Stanford InterServ service



# GIS Viewer Status

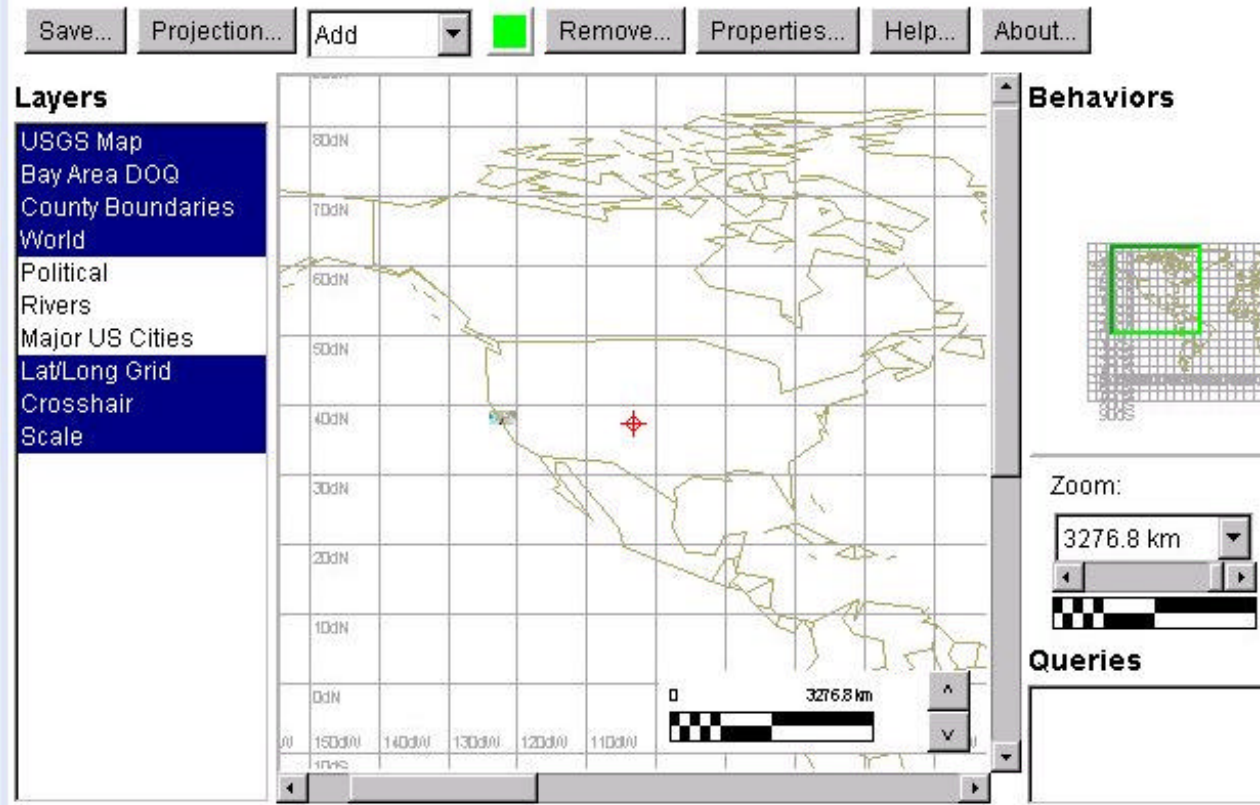
---

- ◆ Layers are geo-referenced data sets
  - Raster: GIF, JPEG
  - Vectors: internal, *ArcInfo Shape files*
  - Grouping format/protocol: tilePix
- ◆ Habanero-ized (with help from UIUC)
- ◆ Support for multiple projections with automatic conversion
- ◆ To be used by Microsoft's next Terraserver release.



# A distant view

This applet requires Java 1.1.5 to work. The following browsers are known to work with it: Microsoft Internet Explorer 4.0 (later releases), Netscape Navigator 4.07 or higher, HotJava 1.1.4. Earlier versions will not work. If you are using IE 4.0 and it does not work, try downloading the latest release from Microsoft.



*\*\*Please be patient while this Java applet loads.\*\**



# Closer - with streetfinder

**Bay Area Street Index**  
[Brdg](#) [Exwy](#) [Frwy](#)  
[Hwy](#) [Pkwy](#) [A](#) [B](#) [C](#)  
[D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#)  
[M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#)  
[U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) [1-9](#)  
  
[Teigland Road](#)  
[Teilh Dr](#)  
[Tekman Dr](#)  
[Telegraph Ave](#)  
[Telegraph Dr](#)  
[Telegraph Pl](#)  
[Telegraph Hill](#)  
[Telegraph Hill Blvd](#)  
[Telfer Ave](#)  
[Telford Ave](#)  
[Telles Lane](#)  
[Temescal Cir](#)  
[Temescal Ter](#)  
[Temescal Way](#)  
[Temescal Creek](#)  
[Tempe Ct](#)  
[Tempest](#)  
[Templar Pl](#)  
[Temple Ct](#)  
[Temple Dr](#)

**1. Enter street address:**

**2. Select search area:**

**Layers**  
[Detailed Map](#)  
[Aerial Photos](#)  
[Bay Area Streets](#)  
[UTM Grid](#)  
[Crosshair](#)  
[Telegraph Hill Blvd](#)

**Behaviors**  
  
Zoom:

**Queries**





# We're there (with photo!)

1. Enter street address: Telegraph Hill Blvd

2. Select search area: Bay Area

**SEARCH**

Save... Add [dropdown] [green square] Remove... Properties... Views... Help...

**Layers**

- Detailed Map
- Aerial Photos
- Bay Area Streets
- UTM Grid
- Crosshair
- Telegraph Hill Blvd

**Behaviors**

Telegraph Hill

Zoom: 200.0 m

**Queries**

1. Enter street address: Telegraph Hill Blvd

2. Select search area: Bay Area

**SEARCH**

Save... Add [dropdown] [green square] Remove... Properties... Views... Help...

**Layers**

- Detailed Map
- Aerial Photos
- Bay Area Streets
- UTM Grid
- Crosshair
- Telegraph Hill Blvd

**Behaviors**

Telegraph Hill

Zoom: 100.0 m

**Queries**



# Tools for Information Management and Collaboration

---

- ◆ “Multivalent Documents” as a primary enabling technology.
  - “Anytime, Anywhere, Any Type, Every Way User-Improvable Digital Document Platform”
  - A framework of *behaviors* and *layers* for extending document functionality.
  - Conducive to developing a “digital library”-centric browser.



# MVD Status

---

- ◆ Implemented behaviors:
  - Media Adaptor: OCR, *full-fledged HTML*, ASCII
  - Search with search hit visualization
  - Structural: alt. select-and-paste, Notemarks
  - Span: hyperlink, highlight, copyeditor marks
  - Lens: OCR, magnify, notes, *Pilot notes*
  - Manager: lens coordination, user interface
  - Other: Ink
- ◆ Habanero-ized (with help from UIUC)
- ◆ Initial experimentation with behaviors with temporal extent





# Searching a page image

File Edit Go Bookmark Style Lens Tool Select Anno CopyEd View Meta Help

Tue Feb 13 19:51:36 PST 1996

VOL. 70 NO. 4

VARIAN: A MODEL OF SALES

653

## II. The Analysis

The maximum number of customers a store can get is  $I + U$ . Let  $p^* = c(I + U)/(I + U)$  be the average cost associated with this number of customers.

PROPOSITION 1:  $f(p) = 0$  for  $p > r$  or  $p < p^*$ .

### PROOF:

No price above the reservation price will be charged since there is zero demand at any such price. No price less than  $p^*$  will be charged since only negative profits can result from such a price.

If a deviant store charged a slightly lower price,  $p - \epsilon$ , with the same probability with which the other stores charged  $p$ , it would lose profits on order  $\epsilon$ , but gain a fixed positive amount of profits when the other stores tied. Thus for small  $\epsilon$  its profits would be positive, contradicting the assumption of equilibrium.

Let us proceed to a detailed formulation of this argument. First note that  $p^*$  can never be charged with positive probability, for when  $p^*$  is the lowest price charged, profits are zero, and if there is a tie at  $p^*$ , profits are negative. Suppose then that  $p > p^*$  is charged with positive probability.

The number of points of positive mass in the probability distribution must be countable. We can find an arbitrarily small  $\epsilon$  such that  $p - \epsilon$  is charged with probability 0. What happens if we charge  $p - \epsilon$  with probability with which we used to charge  $p$  with probability 0. The increase in profits will be

$$Pr(P_i > p - \epsilon \text{ all } i, P_i \neq p \text{ any } i)$$

$$((p - \epsilon)(I + U) - c(I + U))$$

$$- Pr(P_i > p \text{ all } i) (p(I + U) - c(I + U))$$

$$+ Pr(p_i < p - \epsilon \text{ some } i) ((p - \epsilon)U - c(U))$$

Search

profits

Search Clear ☒ Inc Close

Suppose that all stores were charging a single price  $p$  with  $r \geq p > p^*$ . Then a slight cut in price by one of the stores would capture all of the informed market, and thus make a positive profit. If all stores were charging  $p^*$ , each would get an equal share of the market and thus be making negative profits.



# Highlighting and annotating

File Edit Go Bookmark Style Lens Tool Select Anno CopyEd View Meta Help

Parts of image not recognized by OCR not shown; use View/Full Image Only to see. Tue Feb 13 19:51:36 PST 1996

VOL. 70 NO. 4 VARIAN: A MODEL OF SALES 653

## II. The Analysis

The maximum number of customers a store can get is  $I+U$ . Let  $p^* = c(I+U)/(I+U)$  be the average cost associated with this number of customers.

PROPOSITION 1:  $f(p) = 0$  for  $p > r$  or  $p < p^*$ .

PROOF:

No price above the reservation price will be charged since there is zero demand at any such price. No price less than  $p^*$  will be charged since only negative profits can result from such a price.

PROPOSITION 2: There is no symmetric equilibrium where all stores charge the same price.

PROOF:

Suppose that all stores were charging a single price  $p$  with  $r \geq p > p^*$ . Then a slight cut in price by one of the market stores would capture all of the informed market, and thus

If a deviant store charged a slightly lower price,  $p - \epsilon$ , with the same probability with which the other stores charged  $p$ , it would lose profits on order  $\epsilon$ , but gain a fixed positive amount of profits when the other stores tied. Thus for small  $\epsilon$  its profits would be positive, contradicting the assumption of equilibrium.

Note

Hi Hal,  
Nice Paper!  
How about calling these atoms?  
RW

REPLACE WITH: atoms

The number of points-of-positive mass in any probability distribution must be countable so we can find an arbitrarily small  $\epsilon$  such that  $p - \epsilon$  is charged with probability 0. Consider what happens if we charge  $p - \epsilon$  with the probability with which we used to charge  $p$ , and charge  $p$  with probability 0. The increase in profits will be

$$Pr(P_i > p - \epsilon \text{ all } i, P_i \neq p \text{ any } i)$$






Saved annotations on DARPA home page persist and stay in the right place, despite changes to the page.

File Edit Style Lens Tool Select Anno CopyEd View Meta Help

<=



# Defense Advanced Research Projects Agency

## Defense Advanced Research Projects Agency

---

The Defense Advanced Research Projects Agency (DARPA) is the central research and development organization for the [Department of Defense \(DoD\)](#). It manages and directs selected basic and applied research and development projects for DoD, and pursues research and technology where risk and payoff are both very high and where success may provide dramatic advances for traditional military roles and missions and dual-use applications.

This website is divided into

- [Mission and Organization](#) - to find out about a...
- [A description](#) of the Agency's sponsoring focus to help...
- [A list of](#)...
- [Doing Business](#) with DARPA
  - [Information](#)
  - [DARPA](#)
  - [News Releases](#)

*Back to DARPA again, I see!*

*REPLACE WITH: DARPA*

*Italicize region*

*Italicize region*

Note

Hi Ron,

Here are some comments on the DARPA home page (not a copy).

All but the first comment are executable. Double click on one and see.

Toggle "View/Executive Summary" to see what the applet thinks is important.

Here is a hyperlink in a note to the end of this document, where I highlighted some text.



# MVD Plans

---

- ◆ Support project goals by providing
  - “Media adaptors” for common document formats (esp. XML, LaTeX/DVI, PDF)
  - Support for (non-textual) data types having temporal and geographic extent, involving dynamic elements, and data set elements.
  - Improved annotation tools
  - Mechanisms for manipulating multiple annotations
  - Annotation server support.
  - Behaviors to support multi-lingual interactions



# What We Need to Provide

---

- ◆ Technology
- ◆ Some testbed applications
  - to engage and support real users
- ◆ Evaluation
- ◆ Economic analysis to understand the space of possible viable models



# Economics of New Models of Information Use

---

- ◆ Model economic implications of different forms of cost recovery
  - Some salient issues:
    - » who pays (author, reader, institution)
    - » how they are charged (usage-based, flat fee)
    - » how material is vetted (technical reports, conference proceedings, peer reviewed)
    - » degree of aggregation (articles, journals, collections),
    - » terms and conditions of use (individual view only, institution-wide license, unlimited redistribution).
- ◆ Investigate practical details of applying second-degree price discrimination in the context of information goods
  - Explore ways to determine how different classes of users value different services
    - » needed to determine effective pricing of scholarly goods
    - » via marketing experiments and econometric analysis



# Evaluation and Evaluation Methodology

---

- ◆ Assess usability of tools and content analysis.
- ◆ Improve understanding of the work/technology relationship and collaboration.
  - Use ethnographic (and other) methods to study a community's information activities and practices, the use of information technology to coordinate work across space and time, and the impact of the InterLib.
  - Study how users continue to design technology in the process of use, and modify practices to suit technology.
- ◆ Develop better methods of collecting and interpreting data and evaluating DLs suitable for a large and diverse user community.
- ◆ Automated evaluation of information-centric user interfaces



# Evaluating Technology and Testbed

---

- ◆ To be carried out in the LE, among InterBiome users.
- ◆ Measured by
  - Collecting and analyzing data on use
  - Lab-based usability assessment
  - Measuring impact on work
  - Using ethnographic methods to study the information activities and practices, and artifacts used and produced by InterLib Community.
  - Assessing introduction of proposed technology into other environments







# Testbed Activities

---

- ◆ Special Collections Development
- ◆ A Library-based Learning Environment
- ◆ Application of technology to other environments
- ◆ Experimenting with a Fully Electronic System



# Rethinking Information Use

---

- ◆ Support:
  - entire “information cycle”: creation, dissemination and collaboration
    - » organization, access, presentation and preservation, too
  - non-textual material (photos, video, maps)
    - » as well as text-based content
  - primary data sources, informal “publication”
    - » as well as traditional archival product
  - radically new modes of use
- ◆ Scholarly information use is a great start point



# Scholarly Information

---

- ◆ Cost of maintaining collections is growing ~15%/year.
  - The library subsidizes a University's access to world's content.
- ◆ Imposes large delays
  - filters first, publishes later.
- ◆ Digitization?
  - *increased* costs for little additional functionality
  - finding things is still hard, especially non-textual items



# Proposed space of models

- ◆ Each community pays for dissemination of its content.
  - Cost grows linearly with output.
- ◆ Immediate
  - “publish” first, filter later
- ◆ New functionality
  - collaboration, annotation, electronic notebooks supporting all document types
- ◆ Supports all forms of context
  - scientific data sets, images, maps, video, ...



# Document Analysis

---

- ◆ Rationalize and extend DID (document image decoding) implementation
- ◆ Test via a significant conversion task (e.g., a rare botanical manual)
- ◆ Extend DID to tabular layouts using ideas from turbo coding
- ◆ Extend parsing technology to recognition of 2D expressions (e.g., math)



# Text Analysis

---

- ◆ Word-sense-based text retrieval
  - Use newly available resources for word-sense-based retrieval and thematic-relation-based retrieval.
- ◆ Text Data Mining
  - Tools to aid in the automated discovery of useful, information
- ◆ New statistical models for document selection
  - Improvement over LSI



# System Support

---

- ◆ The InterLib Security System
- ◆ Distributed Information Architecture
- ◆ Distributed Geographic Information Use
- ◆ Scaling technology
  - On Line Data Query Control
  - Indexing Multimedia
- ◆ Conforming Repository Standards





# Security System

---

- ◆ Toolkits for building access control systems
  - support set of primitives for authentication, access control, and structuring complex protocols
- ◆ User-interfaces for access control
  - use *scaffolded systems* to support learning of access control mechanisms
- ◆ Safe summary monitoring
  - usage through the use of *secure fissioned data*



# Tools for Information Management and Collaboration

---

- ◆ Electronic Research Notebook
  - integrate digital library services into notebook architecture
  - built on top of MVD
- ◆ Interfaces for supporting complex information seeking/analysis processes
  - Integration of browsing, search, & use
  - Aiding source selection
  - Effective exploitation of metadata



# Cha-Cha Search Engine

SEARCH RESULTS BY: **Cha-Cha**

1-20 of 561 matches [List View](#) [Next](#)

**Table of Contents**

[Berkeley CS Division Home Page](#)

- ▼ [CS Classes](#)
  - ▼ [CS Class Seminar Home Pages](#)
    - ▼ [CS 298-13: \*\*Digital\*\* Information Systems Seminar](#)
      - [Digital Library Architecture](#)
      - ▼ [CS 298-13: \*\*Digital\*\* Information Systems Seminar](#)
        - [Intellectual Property Rights for \*\*Digital\*\* Library Systems](#)
      - ▼ [CS 298-13: \*\*Digital\*\* Information Systems Seminar](#)
        - [The UC Berkeley \*\*Digital\*\* Library Project](#)
        - [Digital Libraries Initiatives at the World Conservation ...](#)
        - ...
        - [The UC Berkeley \*\*Digital\*\* Library Project: What ...](#)
- ▼ [UCB CS Alphabetical Homepage List](#)
  - ▼ [Taku Tokuyasus Homepage](#)
    - [Bookmarks](#)

[UC Berkeley \*\*Digital\*\* Library Project](#)

- ▼ [Information about the Berkeley \*\*Digital\*\* Library Project](#)
  - [Testbed Development for the Berkeley \*\*Digital\*\* Library ...](#)
  - [Document Processing for the \*\*Digital\*\* Library Project](#)
- ▼ [DLI98 Home Page](#)
  - [Attendees](#)
- ▼ [Digital Library Project Tours](#)
  - ▼ [Tour: Berkeley \*\*Digital\*\* Library Images](#)



# A Distributed Information Architecture

---

- ◆ Divide IR functionality into distributed architecture
  - functionality currently represented by Cheshire II
  - to be implemented as CORBA/ILU/Jylu/InfoBus distributed objects
- ◆ Construct middleware services
  - provide resource discovery and IR
  - based on mapping of natural language to entry vocabularies of remote databases
  - Access via MVD (among other methods)
    - » E.g., bibliography automatically becomes hyperlink that attempt to find resource.



# Scaling Technologies

---

## ◆ Indexing Multimedia

- apply and extend GiST et al. to indexing non-alphanumeric data types generated by InterLib content analysis mechanisms.
  - » E.g., “Blobworld” image signatures
- Explore indexing strategies that can exploit the image content analysis types proposed for InterLib.

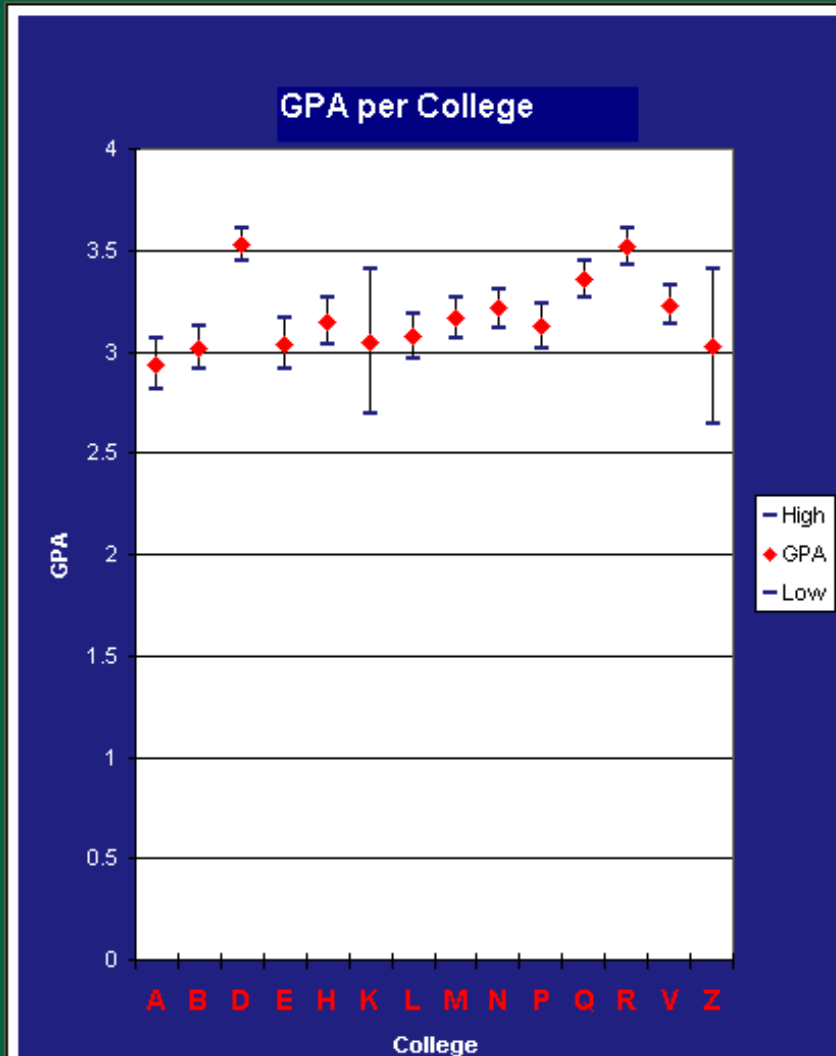
## ◆ On Line Data Query Control

- apply CONTROL technologies to data intensive, long-running of InterLib service components.
- experiment with techniques to allow users to interact with and control continuous feeds,
- examine providing on-line control over large-scale InterLib data extraction tasks that acquire data from multiple InterLib repositories.

# Online Aggregation

Online Aggregation Demo

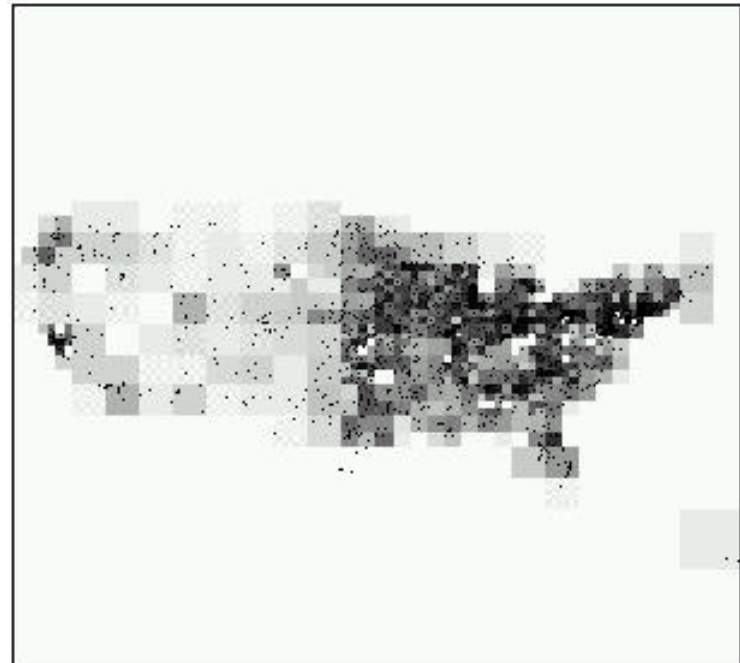
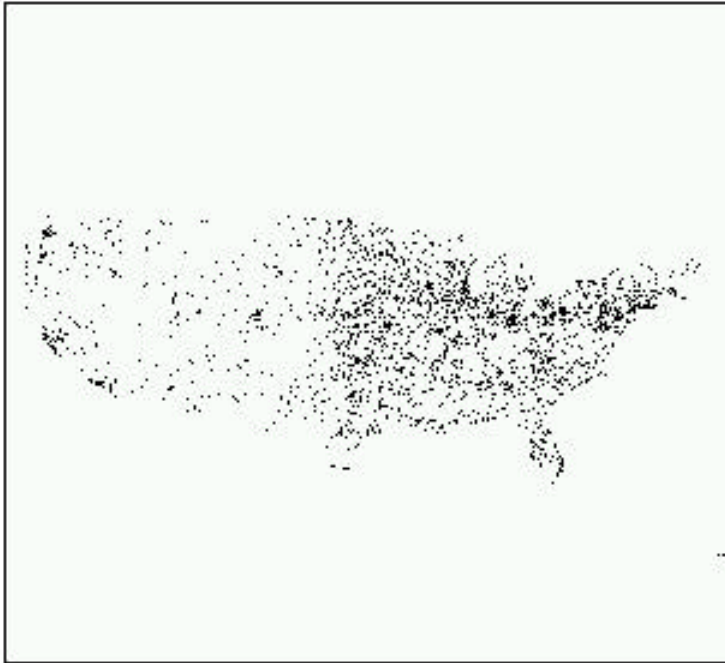
College	Enrolled	GPA
A	190	2.9353933
B	208	3.0174129
D	186	3.5308642
E	167	3.0384614
H	177	3.1497006
K	191	3.0500000
L	185	3.0730338
M	209	3.1683416
N	175	3.2116563
P	177	3.1225805
Q	169	3.3609271
R	184	3.5167599
V	183	3.2289157
Z	22	3.0227273





# CLOUDS

---







# Conforming Repository Standards

---

## ◆ Problem:

- Distributed components having varying availability, stability, persistence attributes.

## ◆ Proposed Solution:

- Define specifications for well-behaved repository.



# A Library-based *Learning Environment* (LE)

---

- ◆ LE allows scholars and students to access InterLib resources and special collections.
- ◆ Collects data use.
- ◆ Provides
  - Discovery Services
  - Research and Collaboration Tools
  - Evaluation and Collaborative Filtering Services
  - “Just-in Time” Mediation